

FILED
CLERK'S OFFICE, CLERK
11 JUL 01 PM 3:54
DISTRICT OF UTAH
BY: _____
DEPUTY CLERK

STUART E. SCHIFFER
Acting Assistant Attorney General
SANDRA M. SCHRAIBMAN
CAROL FEDERIGHI
AMY M. ALLEN
Attorneys, U.S. Department of Justice
Civil Division, Federal Programs Branch
P.O. Box 883, 901 E Street, N.W.
Washington, D.C. 20044
Telephone: (202) 514-4523
Facsimile: (202) 616-8202
Attorneys for Defendants

**IN THE UNITED STATES DISTRICT COURT
DISTRICT OF UTAH, CENTRAL DIVISION**

STATE OF UTAH, et al.,)
Plaintiffs,) Civil No. 2:01-CV-000292G
v.)
DONALD L. EVANS, United States Secretary) DECLARATION OF JOSEPH
of Commerce, et al.,) WAKSBERG
Defendants.)

)

I, Joseph Waksberg, do declare the following:

1. I am currently chairman of the Board of Westat, Inc., one of the largest private firms in the U.S. concentrating on statistical research, survey data collection, and related technology. A major component of Westat's work is in the design of sample surveys and the collection of data. Westat also conducts research on complex sample designs to determine the most efficient sampling methods to use for particular surveys. This research has frequently led to important advances in sampling theory



and methods, and the publication of this research has resulted in the widespread use by sample survey practitioners of new methods.

2. I have been Chairman of the Board of Westat for about 10 years. During most of this period, I was also Senior Vice President attached to the statistical staff. In that position I had a major role in the research and development of sampling methods discussed in paragraph 1. I have been associated with Westat since 1973. Prior to my becoming Chairman of the Board in 1991, I was a Senior Vice President and member of the statistical staff. I retired as an employee in 1999, but remained as Chairman of the Board. At present, I am also a statistical consultant, mostly working with the Westat statistical staff but occasionally for other organizations.
3. From 1940 to 1973, I worked for the Bureau of the Census in a variety of positions. In the last few years of this period I was Associate Director for Statistical Standards, Methodology and Research, and directly responsible for all of the Bureau's research in statistical methodology and the validity and efficiency of sampling methods. I also personally directed and participated in much of the Bureau's research on sample survey theory and methods.
4. I have authored or co-authored many scientific publications and prepared a great many papers for presentation at scientific societies (mostly the American Statistical Association) which have subsequently been printed in the Societies' Proceedings. I have been an active participant in activities of the national and international statistical communities. I am a Fellow of the American Statistical Association, have been on its Board of Directors, and at different times Chair of the Survey Methods Section and the Social Statistics Section of that Association, as well as a member of various Association committees. I am a past President of the Washington Statistical Society. I am a member of the International Statistical Institute (ISI) and the International Association of Survey Statistics (IASS) and was elected a member of the Council of the IASS when it was established. I have been on various panels of the National Academy of Sciences for studies with important statistical components.

I am currently an associate editor of Survey Methodology, a journal devoted to advances in the theory and practice of survey methods.

5. The Department of Justice has asked me to review the Memorandum in Support of Motion for Summary Judgment prepared by the plaintiffs and accompanying declarations by Dr. Lara J. Wolfson and Dr. Donald B. Rubin, and to render an opinion regarding whether the Census Bureau's use of imputation as compensation for missing information in the conduct of the 2000 census constitutes sampling within the meaning of the Census Act, 13 U.S.C. § 195. In my opinion it does not. My reasons are explained in the paragraphs that follow.
6. The particular form of imputation used by the Census for addresses in which no or insufficient data were available does not involve any form of sampling. Dr. Wolfson's declaration includes a definition of sampling, as given in a Dictionary of Epidemiology, John M. Last Ed., 3rd Ed. Oxford University Press, 1995. Sampling is defined as "the process of selecting a number of subjects (units) in a particular group or universe." Dr. Leslie Kish in an authoritative textbook on sampling states "Survey sampling or population sampling, deals with methods for selecting and observing a part (sample) of the population in order to make inference about the whole population." ("Survey Sampling" by Leslie Kish, page 18, John Wiley and Sons, Inc., 1967). The process of selecting one or more units from a larger set of broadly similar units is implicit in these definitions.
The hot-deck imputation used in the 2000 Census was a completely deterministic procedure which utilized data from a predetermined neighbor. There was no process of selecting from among a set of similar units, and, as a result, no sampling.
7. Much of the professional work I have done in the past 40 to 50 years has involved attempting to develop new and better ways of conducting censuses and surveys, evaluating the accuracy of these methods, and comparing them to the efficiency of older methodology. This research includes issues that relate to sampling theory and its applications, and to issues referred to by most textbooks and

research reports as “nonsampling” problems of survey methodology. Most survey researchers and practitioners consider nonresponse and methods of adjusting for nonresponse either by imputation, as it has been done in recent decennial censuses, or through weighting, which is applied more commonly in sample surveys, as nonsampling issues and treat them as separate and distinct from sampling issues.

8. The scientific literature on survey methodology is mostly quite clear on the distinction between sampling and nonsampling issues involved in the design and operations of surveys. (Most researchers use the term “survey” to describe the context of their research because surveys are the dominant form of data collection. However, except for the usual omission of sampling, the measurement problems in censuses are identical to those of surveys, and the problems and methods of dealing with them are the same.) This is illustrated by the references below.

- In 1980, the American Statistical Association printed a short pamphlet entitled “What is a Survey.” (I was one of the co-authors). The publication was intended as a brief guide for nonspecialists on the nature of survey operations and how to judge the quality of a survey. In describing potential sources of errors in surveys, page 17 contains a section entitled “Sampling errors” and page 18 has a section “Nonsampling errors.” One of the examples of nonsampling errors given is “Noninterviews.” The use of imputation in the census, which is involved in this litigation, is solely to reduce the effect of noninterviews on statistics, that is, to reduce nonresponse error, and thus falls into the category of nonsampling methods. The reference is to people in the sample, but the language applies as well to a census.
- The Committee on National Statistics has published a three-volume report “Incomplete Data in Sample Surveys” (Academic Press, 1983). Unit nonresponse, which causes one of the major types of incomplete data, is distinct from sampling error. The first sentence of Chapter 1 states: “Survey results are affected by errors arising from several sources, the most important of which

are sampling error, error due to incomplete data, response error, and processing error."

Similarly, chapter 5 which is a "Review of Theory" states on page 71: "Part V describes a general structural model for survey errors, which include sampling errors, response errors and nonresponse errors, and interactions among those sources of errors." Again, nonresponse and adjustments for their effects are described as methods that are distinct from the sampling operations.

- "Nonsampling Errors in Surveys" by Judith T. Lessler and William D. Kolsbeck (John Wiley and Sons, Inc., 1992), as the title indicates, describes current theory and practices related to nonsampling issues in surveys. Three chapters deal with problems of nonresponse and methods of compensating for missing data. The discussion of imputation takes up twenty pages of one of these chapters.
 - Many other examples could be cited. In a 1987 conference on Telephone Survey Methodology co-sponsored by the American Association for Public Opinion Research, the American Statistical Association, and the International Association of Survey Statisticians, the Sessions were grouped by the kinds of issues discussed. One group was "Sampling for Telephone Surveys"; another group was "Nonresponse in Telephone Surveys." Similarly, there have been major statistical conferences devoted to nonsampling problems in surveys and censuses that included discussions of imputation and other methods of compensating for missing data.
9. Paragraphs 7 and 8 above indicate the reasons for my opinion that the common usage by survey statisticians of such expressions as sampling, sample design, sample estimates, etc. does not refer to imputation for missing data, even when the method of imputation involves some auxiliary sample operation (which was not the case in the 2000 census, see paragraph 6). It is useful to note that, from some broad statistical perspectives, every census can be considered a sample operation. For example, in studying the effect interviewers have on the results of a census (or survey), researchers have

considered the choice of the set of interviewers used to collect the data as a random sample of potential interviewers. The assumptions underlying this approach are plausible, and this model has shed useful light on some of the limitations of censuses and surveys. From this point of view, it is not possible to fulfill the requirement to avoid sampling in the production of census data. Obviously, this is not the interpretation of sampling meant by the Census Act. Similarly, methods of data cleaning or quality control are not considered as part of sampling processes in a discussion of sample estimates.

10. I have read Dr. Wolfson's and Dr. Rubin's declarations in which the methodology for imputing persons to addresses with no information is considered as a form of sampling. I believe they have extended the meaning of "sampling" both beyond the definition used by almost all statisticians involved in research on censuses and sample surveys, and in the context of the Census Act. Paragraph 13 of Dr. Rubin's declaration defines sampling as "...obtaining data from a subset of the population...from which estimates are made about characteristics of the entire population." Paragraph 14 of the same declaration states "'The use of sampling in calculating the population for purposes of apportionment'...refers both (1) to the selection of individual units from the relevant population, and (2) to the process of using those units selected to estimate characteristics of the population from which the sample was drawn." Paragraph 15 basically says the same thing. The imputation in the 2000 census did not select a sample of the relevant population, which was the set of addresses with no information. The units used were from an entirely different population, those for which data were available. As mentioned earlier, no sample was selected at all, that is there was no process of selection from a group of similar units in choosing the units that were substituted for those with missing information. Dr. Rubin thus extends the meaning of sampling beyond his own definition. In any case, the language of the Census Act and the reasons for its passage clearly indicate that the references to sampling apply to the way survey statisticians generally use the term, and not to imputation which is part of data cleaning and quality control usually required in large data collection

efforts. Dr. Wolfson similarly enlarges the definition of sampling to include statistical procedures that are not usually considered sampling, and were not identified as such in the Census Act.

11. Both Dr. Wolfson and Dr. Rubin state that better methods for imputation exist than the procedure used in the 2000 census. I have not compared the techniques used in the 2000 Census to alternative methods but whether the form of the hot-deck procedure used in the 2000 Census is less reliable than a method involving random selection, or superior to such a method, is not relevant to the definition of a sampling method. For the reasons described in paragraphs 6 through 10, it seems clear to me that:

(1) the hot deck procedure used in the 2000 Census does not conform to the definition of sampling used by the vast majority of survey statisticians; and (2) imputation is generally considered a nonsampling issue in surveys, and thus does not fall into the category of sampling methods prohibited by the Census Act.

I declare under penalty of perjury that the foregoing is true and correct.

7/10/01
DATE


JOSEPH WAKSBERG

STUART E. SCHIFFER
Acting Assistant Attorney General
SANDRA M. SCHRAIBMAN
CAROL FEDERIGHI
AMY M. ALLEN
Attorneys, U.S. Department of Justice
Civil Division, Federal Programs Branch
P.O. Box 883, 901 E Street, N.W.
Washington, D.C. 20044
Telephone: (202) 514-4523
Facsimile: (202) 616-8202
Attorneys for Defendants

**IN THE UNITED STATES DISTRICT COURT
DISTRICT OF UTAH, CENTRAL DIVISION**

| | | |
|---|---|-----------------------------------|
| STATE OF UTAH, <u>et al.</u> |) | |
| |) | Civil No. 2:01-CV-000292G |
| Plaintiffs, |) | |
| |) | |
| v. |) | |
| |) | EXHIBITS TO DECLARATION OF |
| DONALD L. EVANS, United States Secretary |) | JOSEPH WAKSBERG |
| of Commerce, <u>et al.</u> |) | |
| |) | |
| Defendants. |) | |
| |) | |

INDEX

| <u>Exh.</u> | <u>Document</u> |
|-------------|---|
| A | Curriculum Vitae |
| B | Leslie Kish, <u>Survey Sampling</u> (1967) (page 18) |
| C | Robert Ferber <u>et al.</u> , American Statistical Association, <u>What Is a Survey</u> (n.d.) |
| D | 1 <u>Incomplete Data in Sample Surveys: Report and Case Studies</u> (William G. Madow <u>et al.</u> eds., 1983) (page 71) |
| E | Judith T. Lessler and William D. Kalsbeek, <u>Nonsampling Error in Surveys</u> (1992) (page viii) |

Exhibits/
Attachments
to this document
have not been
scanned.

Please see the
case file.